



the globus project  
www.globus.org

# Globus Replica Management

Bill Allcock, ANL

PPDG Meeting at SLAC 20 Sep 2000



the globus project  
www.globus.org

## Replica Management

- Maintain a mapping between logical names for files and collections and one or more physical locations
- we define a replica to be a "*managed copy of a file*".
  - The replica management system controls where and when copies are created, and provides information about where copies are located. However, the system does *not* make any statements about file consistency. In other words, it is possible for copies to get out of date with respect to one another, if a user chooses to modify a copy.



the globus project  
www.globus.org

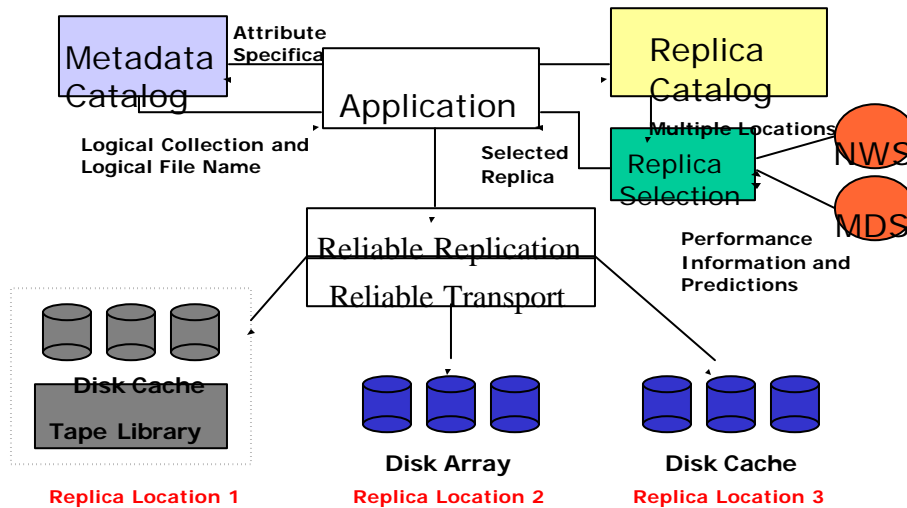
## Our Approach to Replica Management

- Identify replica cataloging and reliable replication as two fundamental services
  - Layer on other Grid services: GSI, transport, information service
  - Use LDAP as catalog format and protocol, for consistency
  - Use as a building block for other tools
- Advantage
  - These services can be used in a wide variety of situations



the globus project  
www.globus.org

## A Model Architecture for Data Grids



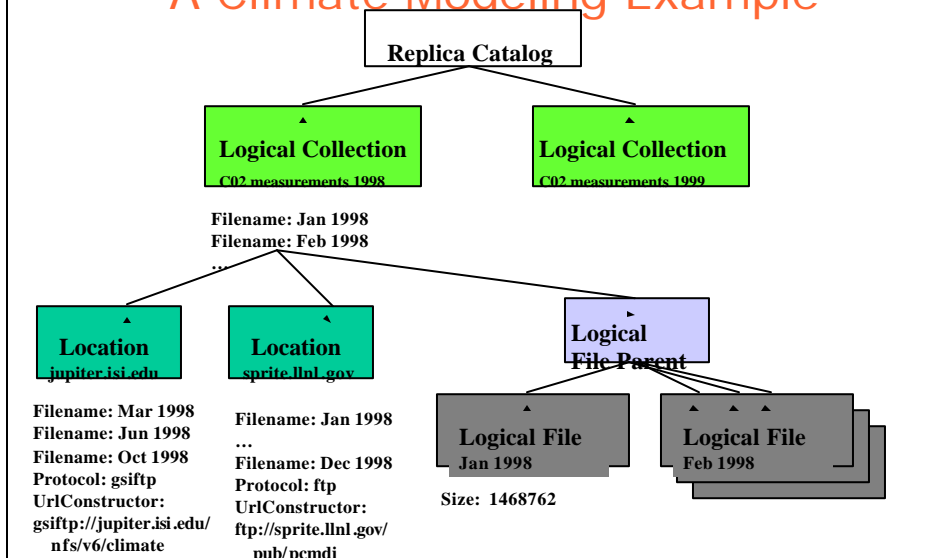


## Replica Manager Components

- Replica catalog definition
  - LDAP object classes for representing logical-to-physical mappings in an LDAP catalog
- Low-level replica catalog API
  - globus\_replica\_catalog library
  - Manipulates replica catalog: add, delete, etc.
- High-level reliable replication API
  - globus\_replica\_manager library
  - Combines calls to file transfer operations and calls to low-level API functions: create, destroy, etc.



## Replica Catalog Structure: A Climate Modeling Example





the globus project  
www.globus.org

## Replica Catalog API

- **globus\_replica\_catalog\_collection\_create()**
  - Create a new logical collection
- **globus\_replica\_catalog\_collection\_open()**
  - Open a connection to an existing collection
- **globus\_replica\_catalog\_location\_create()**
  - Create a new location (replica) of a complete or partial logical collection
- **globus\_replica\_catalog\_collection\_list\_filenames()**
  - List all logical files in a collection
- **globus\_replica\_catalog\_location\_search\_filenames()**
  - Search for the locations (replicas) that contain a copy of all the specified files



the globus project  
www.globus.org

## Replica Management API

- **globus\_replica\_management\_register\_files:**
  - Register a set of files at a source location in a replica catalog.
- **globus\_replica\_management\_copy\_files:**
  - Replicate a set of files from a source location to a destination: i.e., copy the files and update the replica catalog.
- **globus\_replica\_management\_is\_current:**
  - Function that returns a boolean vector that indicates if the specified files are out of date, with respect to a user-provided comparison [function](#). (Note: Just how to implement this function remains to be seen.)
- **globus\_replica\_management\_update\_files:**
  - Update a set of files from a source location to a destination.
- **globus\_replica\_management\_delete\_files:**
  - Delete a set of replicas from a specified location: i.e., delete the files and update the replica catalog.
- **globus\_replica\_management\_synchronize\_filenames()**
  - Ensure that the location object for a physical storage directory correctly reflects the contents of the directory



the globus project  
www.globus.org

## Replica Catalog Services as Building Blocks: Examples

- Combine with information service to build replica selection services
  - E.g. “find best replica” using performance info from NWS and MDS
  - Use of LDAP as common protocol for info and replica services makes this easier
- Combine with application managers to build data distribution services
  - E.g., build new replicas in response to frequent accesses



the globus project  
www.globus.org

## Relationship to Metadata Catalogs

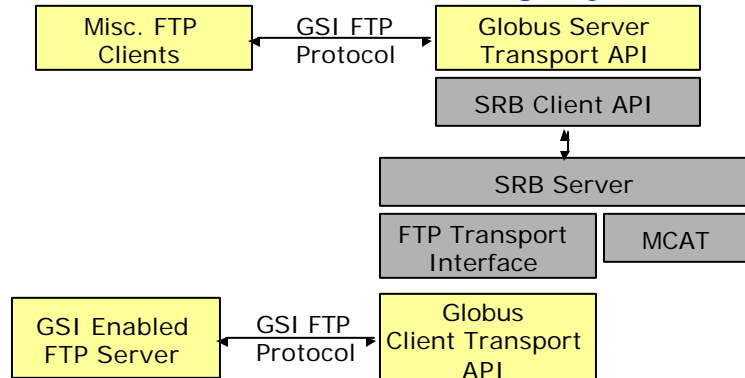
- Metadata services describe data contents
  - Have defined a simple set of object classes
- Must support a variety of metadata catalogs
  - MCAT being one important example
  - Others include LDAP catalogs, HDF
- Community metadata catalogs
  - Agree on set of attributes
  - Produce names needed by replica catalog:
    - > **Logical collection name**
    - > **Logical file name**



the globus project  
www.globus.org

## Globus and SRB: Integration Plan

- FTP access to SRB-managed collections
- SRB access to Grid-enabled storage systems



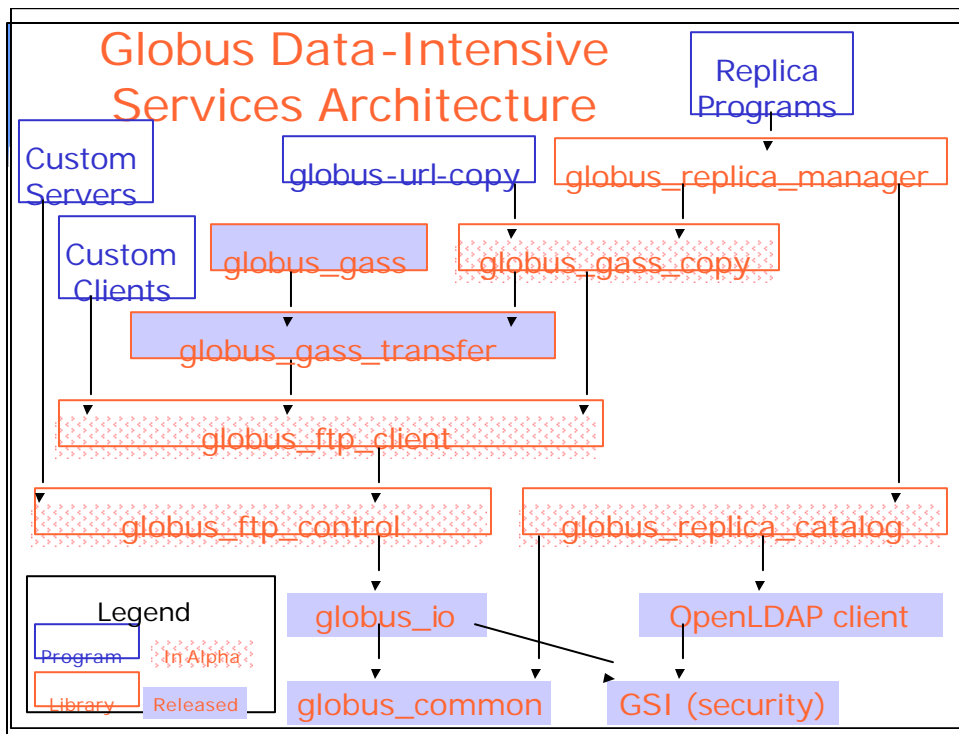
the globus project  
www.globus.org

## Outstanding Issues

- What write consistency should we support?
- Methodology for handling updates
- Access Control
- Replicating the replica catalog
- Replication of partial files
- Alternate catalog views: files belong to more than one logical collection
- Intermediate feedback required (callbacks)
- Timing

## Status

- Grid FTP and catalog management API and tools in alpha test
- Demonstration applications with climate data
- SRB/Globus data grid services integration underway
- Replica Management API under design
- Grid based access control strategy under design





the globus project  
[www.globus.org](http://www.globus.org)

The End